APPENDIX A

TECHNICAL NOTES

TECHNICAL NOTES

General information

The data in this report come from many sources, including surveys conducted by the National Science Foundation (NSF) and other Federal agencies, and by non-Federal organizations. Many methods of data collection are represented, such as universe surveys, sample surveys, and compilations of administrative records. Users should thus take great care when comparing data from different sources. These data often will not be strictly comparable due—among other things—to differences in definitions, survey procedures, and phrasing of questions.

Survey accuracy is determined by the joint effects of "sampling" and "nonsampling" errors. In all of the surveys that are sources of data for this report, efforts are made to minimize these errors. Sampling errors arise because estimates based on a sample will differ from the figures that would have been obtained if a complete census had been taken.

All surveys, whether universe or sample, are also subject to nonsampling errors; these can arise from design, reporting, and processing errors as well as from errors due to faulty response or nonresponse. Nonsampling errors include respondent-based events, such as some respondents interpreting questions differently from other respondents; respondents making estimates rather than giving actual data; and respondents being unable or unwilling to provide complete, correct information. Errors can also arise during the processing of responses, such as recording and keying errors.

Racial/ethnic information

Data collection on and reporting of the race/ethnicity of individuals pose several additional problems. First, both the naming of population subgroups and their definitions often have changed over time. Because this report draws on data from many sources, different terminology may have been used to obtain the various statistics presented here. Efforts have been made to maintain consistency throughout this text, but in some data reporting, it has been necessary to use distinct terminology that does not match that used in other compilations.

Second, many of the groups of particular interest are quite small, so that it is difficult to measure them accurately without universe surveys. In some instances, sample surveys may not have been of sufficient scope to permit calculation of reliable racial/ethnic population estimates; consequently, results are not shown for all groups. The Bureau of the Census's Current Population Survey, for example, cannot provide data on American Indians. Data on this population are available only from the decennial census. Another issue related to race/ethnicity is the fact that it is easy to overlook or minimize heterogeneity within subgroups when only a single statistic is reported for a total racial/ethnic group.

Third, data on race/ethnicity are often based on self-identification. These data are less reliable for certain racial/ethnic groups than for others. Data collected at two points in time indicate that self-identification of American Indians is much less reliable than self-identification of other racial/ethnic groups.[1]

Information about persons with disabilities

Data on persons with disabilities in science and engineering are seriously limited for several reasons. First, the operational definitions of "disability" vary and include a wide range of physical and mental conditions. Different sets of data have used different definitions and thus are

---

[1]U.S. Bureau of Labor Statistics, *A Test of Methods for Collecting Racial and Ethnic Information* (Washington, DC: U.S. Department of Labor, 1995).

not totally comparable. The Americans with Disabilities Act of 1990 (ADA) encouraged progress toward standard definitions. Under ADA, an individual is considered to have a disability if he or she has a physical or mental impairment that substantially limits one or more major life activities, has a record of such impairment, or is regarded as having such an impairment. ADA also contains definitions of specific disabilities.

Second, data about disabilities frequently are not included in comprehensive institutional records (e.g., in registrar records in institutions of higher education). If included at all in institutional records, such information is likely to be kept only in confidential files at an office responsible for providing special services to students. Institutions are unlikely to have information regarding any persons with disabilities who have *not* requested special services. In the case of elementary/secondary school programs receiving funds to provide special education, however, counts for the entire student population identified as having special needs are centrally available.

Third, information on persons with disabilities gathered from surveys is often obtained from self-reported responses. Typically, respondents are asked if they have a disability and to specify what kind of disability it is. Resulting data therefore reflect individual perceptions rather than objective measures.

An example—the attempt to provide estimates of the proportion of the undergraduate student population with disabilities—shows how these factors coalesce. Self-reported data from the undergraduate student population, queried on a survey to ascertain patterns of student financial aid, suggest that about 10 percent of this population report having some disability. Estimates from population surveys of higher education institutions, in contrast, place the estimate much lower, between 1 and 5 percent. Whether this discrepancy is the result of self-perception, incomplete reporting, nonevident disabilities, or differing definitions is difficult to ascertain.

In the final analysis, although considerable information is available on persons with disabilities and their status in the educational system and in the science and engineering workforce, it is often not possible to compare the numbers of persons with disabilities from different sources.

Primary non-NSF sources

The following non-NSF sources were used for data tables in this report.

The Integrated Postsecondary Education Data System Survey: Fall Enrollment, Completions, and Institutional Characteristics

Contact: National Center for Education Statistics
U.S. Department of Education
1990 K Street, NW
Washington, DC 20006
(202) 502-7300

http://nces.ed.gov/ipeds/

The Integrated Postsecondary Education Data System (IPEDS) Survey began in 1986 as a supplement to and replacement for the Higher Education General Information Survey (HEGIS), which began in 1966. HEGIS annually surveyed institutions listed in the current National Center for Education Statistics's (NCES's) *Education Directory of Colleges and Universities*; IPEDS surveys all postsecondary institutions, including universities and colleges and the institutions that offer technical and vocational education. The higher education portion is a census of accredited 2- and 4-year colleges; technical and vocational schools are surveyed on a sample basis.

IPEDS consists of several integrated component surveys that obtain information on types of institutions where postsecondary education is available, student participants, programs offered and completed, and the human and financial resources involved in the delivery of postsecondary education. IPEDS includes surveys of institutional characteristics; fall enrollment, including student age and residence; fall enrollment in occupationally specific programs; completions; finance; staff; salaries of full-time instructional faculty; and academic libraries.

The **IPEDS Institutional Characteristics Survey** provides the basis for the universe of institutions reported in the *Education Directory of Colleges and Universities.* The universe includes institutions that met certain accreditation criteria and offered at least a 1-year program of college-level studies leading toward a degree. Each fall, institutions listed in the previous year's directory are asked to update information on their school's characteristics.

The **IPEDS Completions Survey** replaces and extends the HEGIS Degrees and Other Formal Awards Conferred Survey. It is administered to a census of institutions offering degrees at the bachelor's degree level and above, all 2-year institutions, and a sample of less-than-2-year institutions.

The **IPEDS Fall Enrollment Survey** replaces and extends the previous HEGIS surveys of enrollment in institutions of higher education.

The National Postsecondary Student Aid Study

Contact: National Center for Education Statistics
U.S. Department of Education
1990 K Street, NW
Washington, DC 20006
(202) 502-7300

http://nces.ed.gov/npsas/

The National Postsecondary Student Aid Study (NPSAS) was established by NCES to collect information concerning financial aid allocated to students enrolled in U.S. postsecondary institutions. NPSAS was first administered in the fall of the 1986–87 academic year. NCES conducted subsequent cycles of NPSAS for the 1989–90, 1992–93, and 1995–96 school years. The 1989–90 cycle contained enhancements to the methodology used in the 1987 cycle. Estimates from the 1996 NPSAS sample are generally comparable to those from the 1993 and 1990 samples but not to those from the 1987 sample.

The 1995–96 survey gathered information from about 60,000 undergraduate and graduate students selected from registrar lists of enrollees at about 800 postsecondary institutions. The sample included students who did and did not receive financial aid, as well as students' parents. Student information, such as field of study, educational level, and attendance status (part or full time), was obtained from registrar records. Types and amounts of financial aid and family financial characteristics were abstracted from school financial aid records. Parents of students were also sampled to compile data concerning family composition and parental financial characteristics.

Engineering Workforce Commission Survey of Engineering and Technology Enrollments

Contact: Matt Doster
Engineering Workforce Commission
1111 19th Street, NW
Suite 403
Washington, DC 20036
(202) 296-2237

http://www.aaes.org/ewc/

For 29 years, the Engineering Workforce Commission (EWC) has conducted annual surveys of enrollments in engineering programs. The 1996 report on enrollments in engineering covers 335 institutions including all of those with curricula approved by the Accreditation Board for Engineering and Technology (ABET), as well as data on engineering technology from 285 schools. The response rate to the 1996 survey was 96.1 percent. EWC counts the number of students studying for engineering degrees at all ABET-accredited engineering schools throughout the United States. Historically, EWC has also included schools that are not ABET accredited for a variety of reasons unique to each school. Some such schools are in the process of obtaining ABET accreditation; others have simply asked to be included in the survey. Each year, EWC obtains data from all schools included in the previous year's survey so as to ensure accurate time-series comparisons.

Survey of Income and Program Participation

Contact: Michael McMahon
Bureau of the Census
U.S. Department of Commerce
Washington, DC 20233
(301) 457-3819

http://www.sipp.census.gov/sipp/

The Survey of Income and Program Participation conducted by the Census Bureau provides information on the economic situation of households and persons in the United States. The survey collects data on basic social and demographic characteristics of persons in households, labor force activity, type and amount of income, participation status in various programs, and various supplementary modules, for example, work history, health characteristics (including disability), assets and liabilities, and education and training.

A combined sample from the 1992 and 1993 panels of the Survey of Income and Program Participation provides the latest available data on the disability status of the noninstitutionalized population of the United States. A supplement containing an extensive set of questions about disability status was included as part of the ninth wave of the 1992 panel and the sixth wave of the 1993 panel. Both of these waves were fielded between September and December 1994. The total sample size for this study was approximately 40,000 interviewed households.

The disability supplements that have been asked in SIPP were designed to be consistent with the ADA definition of disability. The supplements obtain information on the ability to perform specific functional activities (seeing, hearing, having one's speech understood, lifting and carrying, climbing stairs, and walking); certain ADLs or activities of daily living (getting around inside the home, getting in and out of a bed or chair, bathing, dressing, eating, and toileting), and certain IADLs or instrumental activities of daily living (going outside the home, keeping track of money and bills, preparing meals, doing housework, and using the telephone). The survey also collects information on the use of such special aids as wheelchairs and canes, the presence of certain conditions related to mental functioning, and the ability to work at a job or business.

People 15 years old and over were identified as having a disability if they met any of the following criteria:

- Used a wheelchair or were a long-term user of a cane, crutches, or a walker
- Had difficulty performing one or more functional activities (seeing, hearing, speaking, lifting/carrying, using stairs, or walking)
- Had difficulty with one or more activities of daily living (the ADLs included getting around inside the home, getting in or out of bed or a chair, bathing, dressing, eating, and toileting)
- Had difficulty with one or more instrumental activities of daily living (the IADLs included going outside the home, keeping track of money and bills, preparing meals, doing light housework, taking prescription medicines in the right amount at the right time, and using the telephone)
- Had one or more specified conditions (a learning disability, mental retardation or another developmental disability, Alzheimer's disease, or some other type of mental or emotional condition)
- Were limited in their ability to do housework
- Were 16 to 67 years old and limited in their ability to work at a job or business
- Were receiving federal benefits based on an inability to work

People age 15 and over were identified as having a severe disability if they were unable to perform one or more functional activities; needed personal assistance with an ADL or IADL; used a wheelchair; were a long-term user of a cane, crutches, or a walker; had a developmental disability or Alzheimer's disease; were unable to do housework; were receiving federal disability benefits; or were 16 to 67 years old and unable to work at a job or business.

Primary NSF/Division of Science Resources Studies (SRS) sources

The following SRS sources were used for data tables in this publication. Published data tables from these surveys may be accessed on the SRS Web page <http://www.nsf.gov/sbe/srs>. In addition, researchers may access data directly from the SESTAT or WebCASPAR database systems, which can be accessed from the SRS Web page.

Survey of Earned Doctorates

The Survey of Earned Doctorates (SED) has been conducted annually since 1957 for the National Science Foundation, the U.S. Department of Education, the National Endowment for the Humanities, the National Institutes of Health, and the U.S. Department of Agriculture. This is a census survey of all recipients of research doctoral degrees such as Ph.D. or D.Sc.; it excludes the recipients of first-professional degrees such as J.D. or M.D. Therefore, SED data are restricted to research doctorates.

Data for the SED are collected directly from individual doctorate recipients contacted through graduate deans at all U.S. universities. The recipients are asked to provide information on the field and specialty of their degree as well as their personal educational history, selected demographic data, and information on their postgraduate work and study plans. Approximately 95 percent of the annual cohort of doctorate recipients respond to the questionnaire.

Partial data from public sources, such as field of study, are added to the file for nonrespondents. No imputations are made, however, for nonresponse for data not available elsewhere, such as race/ethnicity information. The data for a given year include all doctorates awarded in the 12-month period ending on June 30 of that year. Information on the SED can be found on the Web at <http://www.nsf.gov/sbe/srs/ssed/start.htm>.

Survey of Graduate Students and Postdoctorates in Science and Engineering

The data collected in the Survey of Graduate Students and Postdoctorates in Science and Engineering represent national estimates of graduate enrollment and postdoctoral employment at the beginning of the academic year in all academic institutions in the United States that offer

doctorate or master's degree programs in any science or engineering field. Included are data for all branch campuses; affiliated research centers; and separately organized components such as medical or dental schools, schools of nursing, or schools of public health. In fall 1997, the survey universe consisted of 723 reporting units at 601 graduate institutions. Data are collected at the academic department level.

Available information includes full-time graduate students by source and mechanism of support, including data on women and first-year students enrolled full time; part-time graduate students by sex; and citizenship and racial/ethnic background of all graduate students. In addition, detailed data on postdoctorates are available by source of support, sex, and citizenship, including separate data on those holding first-professional doctorates in the health fields; summary information on other doctorate nonfaculty research personnel is also included.

NSF has collected data on graduate science and engineering enrollment and postdoctoral appointees since 1966. From fall 1966 through fall 1971, data from a limited number of doctorate-granting institutions were collected through the NSF Graduate Traineeship Program, which requested data only on those science and engineering fields supported by NSF. Beginning with the fall 1972 survey, this data collection effort was assigned to the Universities and Nonprofit Institutions Studies Group of NSF's Division of Science Resources Studies. It was gradually expanded during the period 1972–75 to include additional science and engineering fields as well as all institutions known to have programs leading to the master's or doctorate degree. Because of this expansion, data for 1974 and earlier years are not strictly comparable with 1975 and later data. Information on the Graduate Student Survey can be found on the Web at <http://www.nsf.gov/sbe/srs/sgss/start.htm>.

NSF's SESTAT data system

In the 1990s, SRS redesigned its data system covering scientists and engineers. Termed SESTAT, the new data system integrates data from three SRS surveys—the Survey of Doctorate Recipients, the National Survey of College Graduates, and the National Survey of Recent College Graduates. The integration of the SESTAT surveys requires complementary sample populations and reference periods, matching survey questions, procedures, and field definitions, as well as weighting adjustments for any overlapping populations.

The surveys provide data on educational background, occupation, employment, and demographic characteristics. These surveys are of individuals and have a combined sample size of about 129,000, representing a population of about 12 million scientists and engineers. SESTAT defines scientists and engineers as those who either received a college degree (bachelor's level or higher) in a science or engineering field or who work as a scientist or engineer. Each of the three surveys that makes up the SESTAT data system collects new data every 2 years. The data reported in this publication were collected in 1997.

SESTAT has as its target population residents of the United States with a baccalaureate degree or higher who, as of the study's reference period, were noninstitutionalized, age 75 or less, and either educated as or working as a scientist or engineer. A baccalaureate-or-higher degree is a bachelor's, master's, doctorate, or professional degree. To meet the scientist or engineer requirement, the U.S. resident had to (1) have at least one baccalaureate-or-higher degree in a science or engineering field or (2) have a baccalaureate-or-higher degree in a non-science or -engineering field but work in a science and engineering occupation as of the survey reference week. For the 1997 SESTAT surveys, the reference period was the week of April 15, 1997.

Some elements of SESTAT's desired target population were not included within the target populations of any of the three SESTAT component surveys. Bachelor's and master's level science and engineering trained personnel missing from the survey frames are predominately:

- residents whose bachelor's and/or master's degrees in science and engineering were received prior to April 1990 or from a foreign institution, who resided outside the United States on April 1, 1990, but not with the U.S. armed forces stationed abroad; or
- residents with no baccalaureate or higher degree in any field as of April 1, 1990, who were awarded a degree in science and engineering after June 1994 by a U.S. institution or after April 1990 by a foreign institution.

Persons with at least a bachelor's degree who are working in science and engineering jobs, but have no degree in a science or engineering field, are underrepresented in the SESTAT database after 1993 because the surveys do not capture new persons entering these occupations who are not educated in science and engineering fields in this decade.

Doctorate-level science and engineering trained personnel missing from the survey frames are predominately:

- residents with doctorates in science and engineering received after June 1996 or from a foreign institution, with no baccalaureate-or-higher degree in any field as of April 1, 1990, and no bachelor's or master's degree in science and engineering received from a U.S. institution between April 1, 1990, and June 1996; or
- residents with doctorates in science and engineering received after June 1996 or from a foreign institution but with no bachelor's or master's science and engineering degree received from a U.S. institution between April 1, 1990, and June 1996, who resided outside the United States on April 1, 1990, but not with the U.S. armed forces stationed abroad.

SESTAT classifies the following broad categories as science and engineering occupations: computer and mathematical scientists, life and related scientists, physical and related scientists, social and related scientists, and engineers. Postsecondary teachers are included within each of these groups. The following are considered non-science and -engineering occupations: top- and mid-level managers; teachers, except science and engineering postsecondary teachers; technicians/technologists, including computer programmers; people in health and related occupations, social services and related occupations, sales and marketing occupations, and other non-science and -engineering occupations—for example, artists, broadcasters, editors, entertainers, public relations specialists, writers, clerical and administrative support personnel, farmers, foresters, fishermen, lawyers, judges, librarians, archivists, curators, actuaries, food service personnel, historians (except science and technology), architects, construction tradespeople, mechanics and repairers, and those involved in precision/production occupations, operators (for example, machine set-up, machine operators and tenders, fabricators, assemblers) and related occupations, transportation/material moving occupations and protective and other service occupations. Information on SESTAT can be found on the Web <http://srsstats.sbe.nsf.gov/>.

Sampling errors

Sampling errors occur when estimates are derived from a sample rather than from the entire population. The sample used for any particular survey is only one of a large number of possible samples of the same size and design that could have been selected. Even if the same questionnaire and instructions were used, the estimates from each sample would differ from the others. This difference, termed sampling error, occurs by chance, and its variability is measured by the standard error associated with a particular estimate.

The standard error of a sample survey estimate measures the precision with which an estimate from one sample approximates the true population value, and thus can be used to construct a confidence interval for a survey parameter to assess the accuracy of the estimate. Standard errors for the numbers in the appendix tables are provided where available. Tables A-1 through A-6 provide standard errors for tables in chapter 1. Tables A-7 through A-10 provide approximate standard errors for totals for different segments of the science and engineering population from the NSF SESTAT surveys. Information provided in tables A-11 through A-14 allows the user to calculate approximate standard errors for estimates derived from the NSF SESTAT surveys. The following formula can be used for estimating the standard error of totals:

$$SE(Y) = [\beta_0 Y^2 + \beta_1]^{1/2}$$

Where: SE(Y) is the predicted standard error of the estimated total Y and β_0 and β_1 are the regression coefficients provided in tables A-11 through A-14. Approximate standard errors for percentages can be calculated from the following formula:

$$SE(P) = [\beta_1/Y (P(100\text{-}P))]^{1/2}$$

Where: SE(P) is the predicted standard error for the percentage, Y is the estimated number of persons in the base of the percentage, and β_1 is the regression coefficients provided in tables A-11 through A-14. A 95 percent confidence interval for an estimate can be calculated by multiplying 1.96 by the standard error of the estimate, and adding and subtracting the resulting amount from the estimate.